Precision and Accuracy

Hartmann

PRECISION AND ACCURACY

BY
GEORGE W. HARTMANN, Ph.D.

ARCHIVES OF PSYCHOLOGY
R. S. WOODWORTH, Editor
No. 100

NEW YORK
May, 1928

ARCHIVES OF PSYCHOLOGY

Columbia University, New York City. The Subscription price is six dollars per volume of about 500 pages. Volume I comprises Nos. 2-10, Volume II, Nos. 11-18, Volume III, Nos. 19-25, Volume IV, Nos. 26-32, Volume V, Nos. 33-39, Volume VI, Nos. 40-46, Vol. VII, Nos. 47-52, Vol. VIII, Nos. 53-58, Vol. IX, Nos. 59-63, Vol. X, Nos. 64-68, Vol. XI, 69-73, Vol. XII, 74-78, Vol. XIII, 79-85, Vol. XIV, 86-91, Vol. XV, 92-98. The available numbers are as follows:

2. On the Functions of the Cerebrum: The Frontal Lobes: Shepherd Ivory Franz. 50c.
3. Empirical Studies in the Theory of Measurement: Edward L. Thorndike. 50c.
4. Rhythm as a Distinguishing Characteristic of Prose Style: Abram Lipsky. 50c.
5. The Field of Distinct Vision: W. C. Ruediger. 70c.
6. The Influence of Bodily Position on Mental Activities: Elmer E. Jones. 50c.
7. A Statistical Study of Literary Merit: Frederic Lyman Wells. 30c.
8. The Relation Between the Magnitude of the Stimulus and the Time of Reaction: Sven Froeberg. 35c.
9. The Perceptual Factors in Reading: Francis Marion Hamilton. 50c.
10. Time in English Verse Rhythm: Warner Brown. 70c.
11. The Hearing of Primitive Peoples: Frank G. Bruner. $1.00.
12. Studies in Development and Learning: Edwin A. Kirkpatrick. $1.00.
13. The Inaccuracy of Movement: H. L. Hollingworth. 80c.
14. A Quantitative Study of Rhythm: Herbert Woodrow. 60c.
15. The Psychology of Efficiency: Henry Alford Ruger. $1.25.
16. On Certain Electrical Processes in the Human Body and their Relation to Emotional Reactions: Frederic Lyman Wells and Alexander Forbes. 40c.
17. The Relative Merit of Advertisements: Edward K. Strong, Jr. $1.00.
18. Attention and Movement in Reaction Time: J. V. Breitwieser. 50c. (Cloth, 75c.)
19. An Empirical Study of Certain Tests for Individual Differences: Mary Theodore Whitley. $1.25. (Cloth, $1.50.)
20. Visual Acuity with Lights of Different Colors and Intensities: David Edgar Rice. 50c. (Cloth, 75c.)
21. The Curve of Forgetting: C. H. Bean. 45c. (Cloth, 70c.)
23. Reaction Time to Retinal Stimulation. A. T. Poffenberger, Jr. 70c. (Cloth, 95c.)
24. Interference and Adaptability: Arthur Jerome Culler. 75c. (Cloth, $1.00.)
25. Reaction to Multiple Stimuli. John Welhoff Todd. 60c. (Cloth, 85c.)
26. A Study in Incidental Memory: Garry C. Myers. $1.00. (Cloth, $1.25.)
27. A Statistical Study of Eminent Women: Cora Sutton Castle. 80c. (Cloth, $1.05.)
28. The Mental Capacity of the American Negro: Marion J. Mayo. 60c. (Cloth, 85c.)
29. Experimental Studies in Judgment: H. L. Hollingworth. $1.25. (Cloth, $1.50.)
30. The Psychological Researches of James McKeen Cattell: A Review by Some of His Pupils. $1.00. (Cloth, $1.25.)
31. Fatigue and Its Effects upon Control: Isaac Emery Ash. 60c. (Cloth, 85c.)
32. The Transfer Effects of Practice in Cancellation Tests: Melvin Albert Martin. 60c. (Cloth, 85c.)
33. The Intellectual Status of Children Who are Public Charges. J. L. Stenquist, E. L. Thorndike and M. R. Trabue. 50c. (Cloth, 75c.)
34. The Relation of Quickness of Learning to Retentiveness: Darwin Oliver Lyon. 50c. (Cloth, 75c.)
35. The Overcoming of Distraction and Other Resistances: John J. B. Morgan. 75c. (Cloth, $1.00.)
36. The Psychology of the Negro—An Experimental Study: George O. Ferguson, Jr. $1.25. (Cloth, $1.50.)
37. The Effect of Distraction on Reaction Time: John E. Evans. $1.00. (Cloth, $1.25.)
38. The Effect of Humidity on Nervousness and on General Efficiency: Lorle Ida Stecher. 90c. (Cloth, $1.15.)
39. The Mechanism of Controlled Association. Mark A. May. 75c. (Cloth, $1.00.)
40. Recitation as a Factor in Memorizing. Arthur I. Gates. $1.00. (Cloth, $1.25.)
41. Mental Fatigue during Continuous Exercise of a Single Function: Thomas Russell Garth. 85c. (Cloth, $1.10.)
42. A Psychological Study of Trade-Mark Infringement: Richard H. Paynter, Jr. 85c. (Cloth, $1.10.)
43. Individual Differences and Family Resemblances in Animal Behavior: Halsey J. Bagg. 70c. (Cloth, $1.00.)
44. Experimental Studies in Recall and Recognition. Edith Mulhall Achilles. 90c. (Cloth, $1.25.)
45. The Morphologic Aspect of Intelligence. Sante Naccarati. 70c.
46. Psychological Examination of College Students: F. Edith Carothers. $1.25.
47. The Effects of Practice on Judgments of Absolute Pitch: Evelyn Gough. $1.25.
48. An Experimental Study of Silent Thinking: Ruth S. Clark. $1.40.
49. Some Emperical Tests in Vocational Selection: Herbert W. Rogers. 75 cents.
50. Adenoids and Diseased Tonsils: Their Effect on General Intelligence. Margaret Cobb. $1.00.
51. An Experimental Study of the Factors and Types of Voluntary Choice: Alfred H. Martin. $1.50.
52. Some Well-known Mental Tests Evaluated and Compared: Dorothy Ruth Morgenthau. 80 cents.
53. Mood in Relation to Performance: Elizabeth T. Sullivan. $1.00.
54. The Influence of Incentive and Punishment upon Reaction Time: Albert M. Johanson. 80c.
55. Psychological Tests Applied to Factory Workers: Emily Thorp Burr. $1.25.
56. A Study of the Relation of Accuracy to Speed: Henry E. Garrett. $1.25.
57. An Experimental Study of Hunger in Its Relation to Activity: Tomi Wada. $1.50.
58. Individual Differences as Effected by Practice: Georgina Stickland Gates $1.00.
59. Studies in Industrial Psychology: Elsie Oschrin Bregman. 90 cents.
60. The Mental Status of Psychoneurotics: Alexander D. Tendler. $1.25.
61. Effects of Attention on the Intensity of Cutaneous Pressure and on Visual Brightness: Sidney M. Newhall. $1.25.
62. The Measurement of Motor Ability: Evelyn Garfiel. 90 cents.

(Continued on inside back cover.)

PRECISION AND ACCURACY

BY
GEORGE W. HARTMANN, Ph.D.

ARCHIVES OF PSYCHOLOGY
R. S. WOODWORTH, Editor
No. 100

NEW YORK
May 1928

ACKNOWLEDGMENTS

The writer wishes to express his appreciation of the assistance rendered by Professor A. T. Poffenberger in outlining the plan of this investigation, and by Professor H. E. Garrett for his painstaking supervision of the final stages of the project. He is indebted to Professor R. S. Woodworth for fruitful suggestions, to Messrs. E. G. Flemming and P. Lecky for aid in securing subjects among members of their classes, and to Professor Gordon Allport of Dartmouth College for the original hint which led to this study.

CONTENTS

Introductory

Conspicuous among the major concerns of present-day psychology is the problem of the nature of mental organization. This term centers in itself a variety of questions such as: What constitutes a single mental function? How does it operate? Over how wide a range of material does a certain ability extend? What is the connection between abilities which introspectively or objectively appear alike or unlike as the case may be?

Couched in its modern form, this problem first received explicit formulation in 1904 in the historic paper of Spearman (22), although its adumbrations were already discernible in the writings of Wolff and later "faculty" psychologists. In this article Spearman presented evidence which seemed to indicate the presence of a "general factor" in test performances. This contention occasioned a controversy which even now remains unsettled. His interpretation did not meet with a very favorable reception in America, and German psychology was not yet sufficiently imbued with the statistical spirit of Betz and Lipmann to foster the idea (see 10, p. 504-518). In France, Binet was occupied with the construction of his first intelligence test, which Spearman claims (22a, p. 60, 66, 71) simply carried into practice his own suggestion of promiscuous pooling for the measurement of general ability, but neglected to make clear the theoretical foundations of the work. The lack of an adequate groundwork of "principles" has unquestionably caused many difficulties, such as those which honeycomb the numerous symposia on the nature of intelligence. It would seem after a generation of polemics that Spearman had somewhat the better of the argument, for recently the "particularists"—to borrow a phrase from continental history—appear to have lost their chief protagonist when Thorndike committed himself to the statement, "One cherishes the hope that some simpler, more unitary fact exists as the cause of intellect and that variations in the magnitude of this fact may provide a single fundamental scale which will account for levels and range and surface." (23, p. 412). A complete summary of the controversial literature with a digest of the laborious re-

searches of the British school, and an outline of his present position will be found in Spearman's latest volume (22a).*

Most of the investigations in this field are foreign to our immediate interest, but there is at least one pertinent theoretical aspect: Is any attribute—speed, memory, honesty, accuracy—a general characteristic of an individual or a particular one? Do these qualities appear uniformly in such a manner that if a person exhibits them in one situation he will also manifest them to a similar degree in other tasks?

The answers to these questions bring us into the field of "group factors," a name which has been employed to differentiate those broad capacities which are intermediate in scope between "general ability" or "general intelligence" and the myriad specific functions acquired in a life-time of learning. If one adopts the view that all the "bonds" involved in learned behavior are specific, then the very concept of a group factor is either mythical or superfluous, unless it be interpreted as the product of an overlapping of common elements. Recent students of personality, however, find this view in conflict with other theoretical considerations. If an individual is simply a bundle of habits without any recognizable systematic interlacing among them, then we cannot say that because a man is neat in the care of his person he will be neat in the care of his room; or, if he is honest in money matters that he will be honest in reporting events. But common speech does single out such broader aspects of behavior and label them. To most understandings, an energetic man is not only forceful in bearing and movement, but highly katabolic in all forms of work. This attitude seems to be represented by Courtis (6) who rejects the specificity theory while admitting the proved fact that each new instance of behavior differs from all others. He states his objection in these words: "It closes the door to discovery. It is as if the early chemists had said, 'Each type of substance is specific. Sugar differs from butter and each from water. Therefore, the concept of elements in untenable.'" As a working hypothesis, he suggests the tentative adoption of the opposite view. "Each instance of behavior is specific. The vast array of specific behaviors, however, seem to have certain characteristics in common. A search must be made for tendencies which seem to operate in different situations, and the isolation of their factors attempted." (6, p. 86.)

* An excellent briefer account is given by Dodd (8).

Precision of definition will be of service here as in most instances, hence we propose to term every quality which appears to be a general characteristic a "trait," using that word as defined by G. W. Allport (1)—"a higher integrated habit-system which represents an individual's characteristic mode of adjustment to environmental activities".

The special problem of this study is to determine whether accuracy conforms to such a definition. Is accuracy a habit of a simple order which remains attached to the specialized behavior which it qualifies, or is it a more highly organized ability represented in diverse functions; and to what extent? A solution of this question is attempted in the following pages.

II. The Problem and Its Background

Before we can determine whether accuracy be an integrated way of behaving, it will be necessary to examine those situations or processes wherein it occurs. Sensory discriminations exhibit the trait in its most elementary form: perception of time-intervals, judgments of weights, estimates of distances. Motor activities—like target-aiming and the coordination demanded in mirror drawing—are the second great field from which illustrations of accurate performance may be drawn. Finally, "intellectual" responses such as following directions and solving simple problems may be analyzed with special reference to this type of behavior.

Since accuracy is found in such a variety of acts it may be useful to attach some uniform meaning to the word. It is proposed here to consider accuracy as that characteristic of measurable behavior which is indicated by relative freedom from errors—an error being considered as a deviation from an objectively correct standard. This definition appears adequate for the present undertaking, as it makes accuracy a definite quality of any test product.

These preliminary remarks have seemed essential to the final phrasing of the subject of this investigation: In a given group, does the proportion of errors which each individual makes remain constant for him (relative to the group distribution) no matter what performance he undertakes?

A need for some such study seems to be apparent from the paucity of experimental work in which this factor is emphasized. Speed has always had its obvious share of attention, and the field of difficulty has recently been carefully analyzed (see Thorndike 23), but the third factor in the numerator of Woodworth's "power" formula (30a, p. 11-14):

$$\text{Power} = \text{Quantity} \times \text{Difficulty} \times \text{Excellence of Result} \div \text{Time}$$

has been relatively neglected. To be sure, the work on handwriting, composition, and other educational scales, appears to indicate the contrary, but a measurement of "excellence of result" in the form of accuracy over a wide range of capacities has never been made the primary task of an investigation.

Since the problem in this form seems to have practically no historical antecedents, despite the fact that wide individual

differences in accuracy are obvious in all test results, it would be well to inquire just what related literature is available. Wells (26) appears to have been the first to point out the possibilities of this field. He constructed a battery with the following test components: 1. Addition (of one constant quantity to twenty two-place numbers) ; 2. Long Division problems; 3. Arrangement of numbers in order of size; 4. Alphabetizing of ten names; 5. Copying nine-place numbers; 6. Letter-square substitution of symbols (20, p. 179). "It is intended that the problems of this test-series shall all be well within the range of the subjects' information. The quality of performance then depends in the first instance upon the *care* with which the tasks are accomplished, not upon special information of correct answers or ability to reason them out. In subjects for whom the series is suited, performance is limited essentially by the subject's *carefulness.* Other test series are limited by knowledge and reasoning ability."

It will be observed that Wells has pointed out a condition which all tests of accuracy must meet—viz., adaptation to a homogeneous group. Since the material employed must be within the range of difficulty of all the subjects, one must have recourse to relatively simple tasks, such as those listed above. When this prerequisite is satisfied, differences in "altitude" cease to be a disturbing factor, even though they cannot be said to be equalized. It removes the possibility of having inaccurate performance by an otherwise "careful" individual, whose errors are primarily attributable to his attempt to work beyond his level.

The early work of Krueger and Spearman (16) employed a technique which has now become conventionalized for studies in "factorial" psychology. To eleven subjects these authors gave five tests—pitch discrimination, spatial thresholds, adding, memory span, and a completion test—and computed the intercorrelations. The now supplanted criterion of the hierarchical arrangement of the coefficients was used to demonstrate the existence of a "general factor" in all these tests. In the case of the adding tests, separate correlations between them when scored for speed and accuracy independently and the other tests were obtained. The authors' justification for this was that they were compelled to disregard "Geschwindigkeit" while studing "Genauigkeit" on account of technical difficulties (wegen technischer Schwierigkeiten). (16, p. 91.)

A later paper of Thorndike, Lay, and Dean (23a) on the "Relation of Accuracy in Sensory Discrimination to General Intelligence" had approached the problem with a different objective. These authors took the A.D. from the standard as a measure of inaccuracy in the case of tests with lines and weights, and correlated the obtained values with intellect as judged by teachers and fellow-students. These measures were simply samples of two mental fields presumably at opposite poles of complexity. A correlation of .23 was interpreted to imply the relative independence of the two functions, and used as a refutation of Spearman's original contention. What would have been of interest was the correlation between the ranks of the subjects for linear discrimination and the magnitude of their thresholds for lifted weights. Unfortunately, this was not presented.

The same general topic was studied in the midst of the "formal discipline" era by Winch (28). Asserting that pedagogs aim to inculcate a *habit* of accuracy in schoolwork in the hope that it will later operate in life-situations, he asked, Will training in accurate numerical computation transfer to arithmetical reasoning? By the use of the method of equivalent groups he came to the conclusion that "It seems possible to improve the accuracy of numerical computation without any certainty that we shall thereby improve the accuracy of arithmetical reasoning." A later study of his (28a) appeared to confirm this view which he phrases as follows: "There exists a high positive correlation of .74 between the two functions, numerical computation and arithmetical reasoning. This high correlation does not appear to involve such a community of function that improvement in one function involves improvement in the other."

Another study whose present relationship is rather indirect was made by Trow (24) on the general factor of confidence. He derived a correlation of .03 for all subjects between achievement scores (correctness of judgment) and confidence scores (subjective certainty) on the same tests. Variability and specificity of the confidence trait were pronounced. "Any test of confidence which does not cover a wide number of situations is futile as an indicator, and its predictions will as likely be wrong as right. It seems possible that this is also true of other character traits. The traits are not constants, but vary with the varying situations."

Among the numerous works on the relation of accuracy to speed, that of Garrett (12) may be mentioned as representative. The existence of an optimum rate of lifting weights was established, and there seemed to be evidence that men who ranked high in one rate tended to hold their relative position in others. A certain narrow type of consistency appears to obtain in this process.

Accuracy has always been considered one of the best indices of mental efficiency and fluctuations in this trait have loomed large in the researches of pharmaceutical psychology. Typical of this field is the study of Hull (14) on the effect of tobacco smoking on the accuracy of cancelling A's and mental addition.

Suggestions as to the technique to be employed in the solution of our group factor type of problem have been derived from various sources, such as Hargreaves' (13) study of the imagination, Bernstein's (3) work on quickness, Davey's (7) on "pictoriality," the dissertation of Brown (4) on caution, and that of Dowd (9) on the consistency of the rate of work. In all these cases a definite attribute of a test product was examined for its degree of "spread" or generality, but the results uniformly indicate that none of these traits are "broad" functions. For example, Davey found that a verbal mental test measures the same general factor "g" as does a test similar in form but non-verbal in material, such as pictures, from which she concludes that there is no group factor of "pictoriality."

The investigations of Bernstein and Dowd are of especial interest because both were unable to discover any evidence for a general speed factor. Bernstein found no significant difference between the correlations of his "leisure" and "haste" tests with a criterion of intelligence; this was confirmed by Dowd who showed that the intercorrelations for tests of rate of work for dissimilar performances were very low.

Positive results in the attempted isolation of new group factors have been achieved by a number of Spearman's pupils. We may mention the work of Carey (5) on the nature of specific mental factors. He was unable to unearth a discriminatory function apart from "g," but he did seem to find evidence for the existence of a small general memory factor. Webb's (25) study is especially noteworthy as he apparently demonstrated the existence of a new general factor of "will" (i.e., persistence of motives). This trait functions in the conative sphere in much the same way that "g" operates in the cognitive

life. The factor of perseveration, or persistence of impressions, appears to have been established by the experiments of Lankes (17), and the elaborate mathematical analysis of Webb's data made by Garnett (11) disclosed a new magnitude termed "cleverness," a trait which is considered to be the obverse of perseveration. An incidental result of all these investigations is that the methodics of determining generality have become pretty well standardized, thereby facilitating comparison of data.

III. The Tests and Procedure

1. *Methodological Requirements.*

To discover whether accuracy be a group factor distinguished by functional unity we must first meet the requirement of having a comparatively large number of tests. The justification of this demand is that no one measure of any attribute is an adequate index of how much an individual possesses of it. This is almost axiomatic and follows from the principle that any measurement of ability is in essence an inventory; consequently, the greater the variety of tasks presented, the greater will be the probability that the obtained magnitudes will be fair representatives of the true ones.

Another desideratum to be fulfilled is that the size of the group employed be large enough to serve as an adequate sample of the particular population from which it is drawn. Failure to work with reasonably large groups is a common defect of many of the early studies in this field, both on the part of Spearman's disciples and his opponents. In this investigation eighty-nine Columbia University undergraduates, all men, were used as subjects for the experiments to be described. They were all of sophomore rank or beyond, students in classes of elementary psychology, and each man's presence in the experimental program was purely voluntary. A number of the records were incomplete or unserviceable because of technical errors, so that the final number was reduced to eighty-one. The testing period for each man was about two and one-half hours, a little over an hour being spent on two separate days.

All the tests were administered individually. This would have been inevitable with some of the tests, but it was decided to extend this procedure to every test, partly for the sake of uniformity and partly to exclude the influence of social stimuli upon work done under competition. This precaution was suggested by the results of experiments by Mayer, Schmidt, Meumann, and Allport, which indicate that "stimuli from the group have a favorable effect upon the amount but not upon the quality of work." (2, p. 269.)

2. *Selection, Description, and Administration of the Tests*

The choice of the tests to be used was the first major difficulty encountered in the course of this investigation. If we

cannot even be certain that a completion test is a measure of reasoning ability, then it is clear that simply calling a test a "test of accuracy" does not make it such. But this does not prohibit an attempt to see whether accuracy defined as relative freedom from errors in different test-situations—regardless of the major ability involved—behaves as a unit factor or not.

More than a score of tests were considered for this purpose. The ones finally selected appeared to the experimenter to give the maximum diversification of material available within the restricted time-limits. Among the tests originally proposed, but ultimately eliminated, were: simple letter cancellation, target-aiming, McDougall dotting test, the tracing board, mirror-drawing, estimates of time-intervals, jumbled paragraphs, the I-T test (see Hargreaves, 13), judgments of handwriting, from a series of pairs of digits marking those pairs where the difference between the digits equalled three, and the Thurstone Clerical Test. The reasons for elimination varied, but the desire to avoid duplication and to insure ease and simplicity of administration were the chief motives. The tests remaining seemed to possess most of the desirable features of the rejected tests, and to offer the possibility of measuring the greatest number of typical abilities in which accuracy plays a part.

Records were obtained from each subject for the eleven tests listed below. The number and order of the tests described in the following paragraphs will be constant throughout this report.

(1) Discrimination of lengths. The Galton bar and accessories were placed in uniform diffuse illumination. The subject was seated with his back to the light, about five feet before the bar, and made his adjustments by turning the attached rod. He was kept in ignorance of the size of the standard being used, and was instructed to make his adjustments rapidly and with as little hesitation as possible. After each adjustment the experimenter recorded the reading to the nearest millimeter on a specially prepared sheet. First a ten-centimeter standard was set to the left of the central division point on the Galton bar; the subject was directed to lay off an equivalent distance to the right. Ten trials were taken from the center outwards and ten in the reverse direction, making twenty readings for this standard. A similar

record was secured with the standard distance set at twenty centimeters.

(2) Discrimination of brightness. The Woodworth photometer containing a four candle-power stimulus light on a sliding lever was employed in a dark room with subjects who had been dark-adapted for about twenty minutes before the series was begun. The subject sat with his eyes on a level with the viewing-tube of the instrument and about two inches from it. The pointer was set at zero, so that the two windows were equally illuminated. The experimenter then began to move the bar slowly in the box; the subject who had been directed to consider the left window as the standard reported as soon as he believed the right window to have become just noticeably darker than the left. The pointer reading in millimeters was recorded under "increasing differences". Beginning at a point where the right side was definitely darker than the left, the motion of the bar was reversed until the subject reported the right window equal to the left. This record was entered under "decreasing differences". Corresponding determinations were made on the opposite side of the zero mark. The entire cycle of four readings was performed five times, making twenty readings in all. A similar procedure was undertaken with an eight candle-power lamp.

(3) Pitch discrimination. The Seashore phonograph record was used in this connection, and administered in accordance with the manual of instructions issued by the manufacturers.

(4) Discrimination of intensity of sound. Here, too, the Seashore record seemed to be appropriate to our purposes.

(5) Discrimination of lifted weights. The usual revolving table weights of the laboratory appeared too cumbersome for this diagnosis, and recourse was had to the lighter cartridge weights and the method suggested by Whipple. His recommended procedure involves the use of an eighty-gram standard and twenty-three heavier comparison weights. The procedure given in his text (27, I, p. 225) was followed to the letter. This method is relatively crude, but our justification for using it is that the determination of an exact difference limen with a minute and elaborate technique is quite a different process from this simple determination of functional capacity for comparative purposes. In our experiment two differential thresholds were obtained independently—one

around the eighty-gram weight as a standard and the other around ninety grams.

(6) Attention. This is a hidden word test taken from W. S. Foster's "Notes for Instructors to Accompany Experiments in Psychology," p. 9-11 (see Appendix I). With a time limit of two minutes, the subject was directed to underline all common English words of two, three, or four letters to be found in paragraphs of pied material. A cross was made at the point reached in the text when the end of the period was announced by the experimenter. An alternate form was given so that the test's reliability would be available.

(7) Recognition. This is a stock memory experiment of the manuals. The subject was shown twenty words printed on cards at the rate of about one item per second; these were then mixed with twenty confusion stimuli, and the shuffled pack returned to the observer to be sorted into two piles—one containing those cards he believes he saw and another containing those he did not see. He was permitted to make a third pile for doubtful cases, but subjects rarely availed themselves of this privilege, since it was explained to them that it was not to their advantage to do so. A second set with different words was given as a check on the first (see Appendix II).

(8) Aussage. This is the standard picture report or testimony experiment recommended by Whipple (27, II, p. 27). The observer was shown the lithograph called the "Australians" for twenty seconds and then handed the questionnaire given on page 29 of the preceding reference. Subjects were told to answer as many items as they could, but freedom to guess or omit answers was accorded. In addition, the lithograph entitled "The Disputed Case," a more difficult picture for the perception of details, was given. Following the presentation, the questionnaire from pages 27-28 of Whipple's text was filled in at the subject's leisure.

(9) Perseveration. Adoption of this test was suggested by Bernstein's (3, p. 12) use of it in another context. It consists of an irregular pattern of lines. For the first trial the subject was required to reproduce as many sections as he could in a minute. The pattern was again returned to the subject, this time with changed instructions as follows: "Copy this pattern in the spaces provided below in accordance with the following scheme: Convert every horizontal line into a vertical, and every vertical into a horizontal, retaining the

obliques unmodified." Obviously interference effects are pronounced in this set, so the time was extended to two minutes. When half of the interval had expired, it was announced by the experimenter and the subject made a mark in his copy. This permitted the calculation of the reliability coefficient.

(10) Wells test-battery. This has already been described on page 9. No further instructions were added to those given on the blank, save that the subjects were told at the outset to work as rapidly and as correctly as possible.

(11) Otis directions test. The first four elements in Test I of the Otis Advanced Group Examination were eliminated as too easy, leaving sixteen items to be employed here. It seemed fair to consider errors in this task as due primarily to carelessness, since Test I of Army Alpha, a similar series, rarely yields any mistakes with persons of college level.

It will be noticed at once that the eleven tests fall into two broad classes: the first five are really measures of sensory discrimination, and as such might be called "precision" tests, because they are an index of the general goodness of the afferent nervous system, showing the limits of its capacity; the last six instead demand higher perceptual processes for their execution, and may be called "accuracy" tests by contrast, as they measure the *usual* or customary level of correctness. The first group of tests shows how accurately a man *can* work; the second group indicates how accurately he *does* work.

3. *Scoring the Tests*

The scoring technique employed in this study was especially selected to accord with the nature of our problem, and is based upon certain assumptions which will demand consideration in the interpretation of the final results. The procedure adopted for each test follows:

(1) In the case of the Galton bar, the crude average error of each individual—i.e., the average deviation of all settings from the true setting—was computed for both standards. Presumably the smaller this value the more precise the person who gave it.

(2) For the photometer the average of all twenty readings yielded the just noticeable difference.

(3) The subject's judgment of pitch was graded according to the key furnished by Seashore. The largest number of cor-

rect responses distinguished the most capable man in this test.

(4) Intensity was scored in the same manner as pitch.

(5) Determination of the limen for lifted weights, as usual in a short series, was indirect. The number right—e.g., 6, 7, 8, and 9—was found for each comparison weight, and the 75% threshold for each computed by means of the Fullerton-Cattell table. These derived values for each comparison weight were then averaged to give the final relative threshold.

(6) The attention test was scored on the basis of the percentage which the correctly underlined words were of the total number attempted. This was easily obtained as the subject had made a cross under the letter he was examining when time was called on him. It was assumed that where this proportion was high, the greater must have been the degree of care exerted to attain that result.

(7) The individual's ability in recognition was obtained by taking twice the number of misplaced cards in each pile and subtracting from the total in both piles. To illustrate: One subject had 22 cards in his "Seen" group and 18 in his "Not Seen" group, with 6 errors in the former and 4 in the latter. His score was 40 — 10 (2) = 20 right; or 50%. The few individuals who made a "Doubtful" pile were ranked by adding twice the number of errors to the doubtful cases, and then subtracting this total from 40.

(8) The "Aussage" tests were scored by finding each person's percentage of right answers out of the total number of questions attempted. This amounts to a disregard of the extent of a man's testimony and to limiting comparison to the fidelity of the report he wishes to give. The maximum score is 20 possible right items for the Australians, and 60 possible for the Disputed Case.

(9) The perseveration test was not marked after the manner of Bernstein (3, p. 12), but for the sake of uniformity in our procedure the per cent right of those attempted was obtained for the second trial only. An error consisted in any line omitted or drawn in a different direction from that required by the model and the instructions.

(10) In the Wells test the amount done was constant for each subject. By considering the two division problems as equivalent to five units each, the total possible number of items becomes 66. The per cent right was also secured here.

(11) The Otis test was another amount-limit one. The key furnished by the author was used, leading to a score in terms of per cent right.

The alternate forms in tests 6, 7, 8, and 10 were combined by averaging the individual's per cent right in both. It is this joint measure which was later employed in the intercorrelation of the tests.

4. *Sources of Error in the Experimental Conditions*

A legitimate objection to the scoring procedure used in tests 6, 9, 10, and 11 would be that it is arbitrary to disregard the speed with which a subject works, because one man may do only ten correct items of twenty attempted and another thirty out of sixty, and yet both receive a 50% accuracy standing. The only way to remove this objection would be to weight the accuracy of the performance by the rate of work. This, however, introduces a total efficiency score, which was considered irrelevant to our purpose.

The first five tests in our series seem to be free from this speed complication. In tests 1 and 2, the rate at which the subjects worked was regulated by their own convenience, and in tests 3, 4, and 5, they simply had to work in rhythm with the stimuli.

The "accuracy" tests, however, are not so easily disposed of, because the scores obtained by the method used in the preceding section are probably affected by both speed and difficulty elements. Theoretically, the difficulty is not the same for every subject, because of differences in practice, interest, etc. However, it is hard to see how this can vary to any great extent with the simple tasks here presented to a selected body of college students. To be sure, it is more difficult for some men to be accurate than others. A discussion of the presumptive effects of differences in intellectual level as such is postponed to Section 7 of Chapter IV.

There is no speed element present in tests 7 and 8; in tests 10 and 11, the subjects under identical instructions chose their own rates of work. The resulting accuracy scores, however, are just as indicative of *carefulness*, even if they are the result of varying speeds. Diagnostically, then, there can be no objection to a comparison made from these figures.

The speed factor present in tests 6 and 9 cannot be ex-

plained away as in the other instances. Special treatment is accorded this topic in Section 8 of Chapter IV.

Certainly the subjects of the experiment were completely unaware of the objective of the testing program. They were informed that this was a study designed to see whether their ability in one test was accompanied by ability in another field, and most of them seemed to be well satisfied with this general explanation. No special set toward either speed or accuracy was given by the instructions preceding each "paper and pencil" test.

It may further be remarked that the retention of the space error in the case of the Galton bar and the photometer in no way affects the course of this problem. Elimination of the space error would have been essential in a study of absolute limens, but where relative thresholds are our objective, as in the present instance, the uniformity of the experimental conditions assures valid comparisons of individuals.

IV. PRESENTATION AND DISCUSSION OF RESULTS

1. *The Means and Standard Deviations*

Table I summarizes the original data for all eleven tests. The numerical values in all cases represent the group's average per cent right of the items attempted, except for tests 1, 2, and 5 which are given in threshold values. The records are complete throughout for eighty-one subjects.

Through the files of the University Admissions office*, access was had to the Thorndike scores of the students with a view to determining the representativeness of the experimental group of the Columbia College population. Unfortunately, only sixty-eight records were available, but these gave an average Thorndike score of 82.15, S. D. 13.29. This is not significantly different from the figures given by Wood (29, p. 51). Apparently, then, the subjects who voluntarily presented themselves for experimentation were a fairly random sample of their class—at least as far as intelligence is concerned.

2. *The Reliability Coefficients of the Tests*

Reliability coefficients for tests 1 to 5, inclusive, were found by correlating the magnitudes of the first half of each test with the values of the second half, and correcting the resulting coefficient by the Spearman-Brown prophecy formula for a

TABLE I

MEANS, STANDARD DEVIATIONS, AND ERRORS OF MEASUREMENT
(in terms of per cent right unless otherwise indicated)

Test	Mean	S.D.	S.D.(M)	S.D.(M) / Average
1. Lengths (in cms.)	9.07	4.91	.0982	.0108
2. Brightnesses (in cms.)	5.16	1.23	.0369	.0072
3. Pitch	77.19	10.82	3.2460	.0421
4. Intensity	89.61	5.75	2.7600	.0308
5. Weights (in gms.)	9.35	3.54	2.7258	.2915
6. Attention	62.56	13.91	6.8159	.1089
7. Recognition	68.12	15.28	9.7792	.1436
8. Aussage	63.92	8.90	7.8320	.1225
9. Perseveration	55.74	20.95	7.7515	.1391
10. Wells	87.01	12.71	8.3886	.0964
11. Otis	79.66	11.38	8.8764	.1114

Tests 1, 2, and 5 are given in threshold values.

* The writer is under obligation to Director A. L. Jones for his assistance in this regard.

test of double length. In tests 6, 7, and 10 the reliability was derived by the administration of two alternate and presumably equivalent forms of the tests. Test 9 yielded its reliability by means of the split-half technique, and in test 11 the odd items were correlated with the even. Test 8—the Aussage—was treated in a slightly different fashion, since it was desired to combine the results on both pictures into a joint measure of fidelity of report. As the questionnaires were of unequal length—the Australians containing only 20 items and the Disputed Case 60—the per cent right in the top half of each was added to the per cent right in the bottom half of the other, from which the self-correlation was obtained. In all instances the Spearman-Brown prophecy formula was used.

The reliability coefficients thus secured will be found in Table III, where they lie along the diagonal of italicized figures. It will be observed that the tests of a psycho-physical nature are in general superior in self-consistency to the more complex perceptual variety. Many of the reliabilities are disappointingly low, particularly the Aussage, the Otis Directions, and the Lifted Weights. The unreliability of the Otis directions test is surprising, in view of the fact that it is an integral part of an accepted standard measure of intelligence. Inspection of the frequency of errors in the different items of that test seemed to indicate that the even elements caused more mistakes than the odd; this hypothesis was tested by correlating pairs of items instead of single items, i.e., items 1-2, 5-6, 9-10, and 13-14 were matched for errors against items 3-4, 7-8, 11-12, and 15-16. Even this procedure failed to raise the reliability coefficient originally obtained.†

The reliabilities of tests 1, 2, and 5 shown in Table III are all for the second or greater standard. The low self-correlation of the weights is regrettable, but to have raised it by adding judgments would not have been feasible as this test is the most time-consuming of all in the program for the subject.

The lowest reliability coefficient attaches to the picture report test, and one is at a loss to account for this. It is conceivable, however, that this test—and some of the others with reliabilities not up to the accepted standard of .80 or .90—may be perfectly valid, but that the subjects' own variability

† A personal communication from Dr. Otis suggests that the homogeneity of our experimental group may be responsible for the low reliability coefficient here presented.

in accuracy of performance is so great as to occur within the compass of a single test-situation. Unfortunately, the two facts cannot be isolated from the reliability coefficient alone.

It is of some importance to know whether the scale of measurement used in this study is fine enough to reveal slight differences in accuracy. The obvious answer to this is that the testing instruments vary in this regard. The usual method is to calculate the standard error of measurement which has been done in column 3 of Table I, but since the tests are scored in different units, it is necessary to compare the ratios obtained by dividing the average into the standard error of measurement. Column 4 of Table I shows that the psychophysical tests are superior in this regard, with the exception of the weights which is the poorest.

3. *Consistency of Thresholds on Two Standards*

Of some interest in connection with our problem of the spread of accuracy between tests is the question: Does an individual who occupies a certain relative position in a series of limens for one psychophysical standard tend to maintain this same level when the standard is changed? The data from tests 1, 2, and 5 may be used to throw some light on the matter. Table II presents in the first column of figures the reliabilities of the first and smaller standards, and in the second column the reliabilities of the second and larger standards.

TABLE II

RELIABILITIES OF TWO LIMENS AND THEIR INTERCORRELATIONS

Tests	*Self r of First Standard*	*Self r of Second Standard*	*Intercorrelations of Both Standards*
1. Lengths	.97	.96	.52 ± .05
2. Brightnesses	.90	.91	.79 ± .03
3. Weights	.25	.41	—.02 ± .07

The smaller of the two weight thresholds has practically no reliability for purposes of argument, but the table as a whole seems to indicate that while there is a tendency toward association, still one cannot predict an individual's performance on a larger standard within a given sense modality from a knowledge of his relative position on a smaller. Thus, it is apparent that because a man is good at judging short distances he will not necessarily be good at estimating longer

ones. It would be interesting to see if further research confirms this suspicion.

4. *The Intercorrelations of the Eleven Tests*

Table III presents the complete series of product-moment correlation coefficients of zero order. Inclusive values for the probable errors are given at the bottom of the table, in order to avoid the confusion resulting from the usual method of distributing the probable errors along with their corresponding coefficients in the body of the table.

The correlations have not been corrected for attenuation, because this procedure is probably not, in general, of very great value. "If the correlates are of low reliability, the corrected coefficients cannot be considered of very great reliability either. On the other hand, if the correlates are very reliable the attenuation correction will necessarily be very small. In any event, the corrected coefficients are merely estimates of what might have been obtained had there been no disturbing factors or errors of measurement, a condition which never obtains in practice." (Lemmon, 14, p. 28-29.)

A glance at the table reveals a striking preponderance of very low correlations. It would seem to follow almost directly that whatever factor or group of factors may be found in all these tests must be present to a minimal degree. Only five of the fifty-five raw intercorrelations are reliably greater than zero.

It is apparent that the discrete sense modalities operating in the psychophysical portion of the test series are practically independent of each other. Even within the same sensory field —viz., vision—a person's standing on the Galton bar is no index of what he will do on the photometer as is shown by the r of .005. Pitch and intensity of sound, on the other hand, overlap to the extent indicated by the r of .54. Perhaps this is really due to musical ability rather than to auditory capacity, if one may be permitted to make a distinction that is not intended to be casuistry.

The correlation of .42 between recognition and brightness discrimination is highly mysterious. One might be tempted to speak of a visual ability extended to include visual memory, had not such a suggestion been deprived of support in the preceding paragraph.

The only remaining correlation which is relatively high is

TABLE III

OBTAINED RAW INTERCORRELATIONS WHEN INDIVIDUAL PERFORMANCES ARE SCORED FOR ACCURACY ONLY

Tests	1	2	3	4	5	6	7	8	9	10	11
1. Lengths	*.96*										
2. Brightnesses	.005	*.91*									
3. Pitch	—.02	.20	*.91*								
4. Intensity	—.001	.12	.54	*.77*							
5. Weights	.27	.02	.36	.37	*.41*						
6. Attention	.07	.02	.08	.16	.19	*.76*					
7. Recognition	—.14	.42	.01	.09	.15	—.03	*.59*				
8. Aussage	.15	.20	.15	.23	.18	.10	.08	*.22*			
9. Perseveration	.03	.25	.21	.22	.04	.12	—.11	—.008	*.86*		
10. Wells	—.10	.02	.17	.05	—.10	.08	—.005	.26	.08	*.56*	
11. Otis	.08	.01	.09	.07	—.07	.11	.23	—.003	—.02	.42	*.39*

The P.E.'s of the correlations range from .07 with r's of .00 to .30 to P.E.'s of .01 or less when the r is .90 and beyond. Roughly, the P.E.'s increase by .01 as the r's increase by .10, from .30 on.

N = 81

that obtaining between the Otis and Wells tests. From the nature of these two it might be suspected that differences in intelligence level are responsible therefor, but the evidence offered by the correlations with intelligence is rather equivocal (see Table IV below).

There seems, then, to be little reason to doubt that, as far as general predictive purposes are concerned, our estimate of an individual's level of accuracy from a knowledge of his performance on one or even several tests will be hardly more than a chance guess.

The results so far obtained seem to be in accordance with the findings of earlier investigators. Kitson (15, p. 51) secured a correlation of —.02 ±.10 between printed hard directions and oral hard directions when scored for accuracy. His subjects were a group of college students similar to the ones employed in this study. It is interesting to note the unusual degree of specificity present in two forms of the same material.

The figures given by Rosenow (20, p. 23-26) tell much the same story. When scored for accuracy, tests of opposites, constant increment, and hard directions yielded intercorrelations of .12, .16, and —.01, respectively.

It should be observed that tables of intercorrelations derived from tests scored for accuracy alone are exceedingly rare in the literature. Occasionally, a small grouping like that of the preceding references may be found, but never a complete series.

5. *Relation of the Tests to Intelligence*

In Table IV will be found the correlations of the eleven

TABLE IV

CORRELATIONS OF TESTS WITH THORNDIKE SCORES

Tests	*r with Thorndike*	*P.E.*$_r$
1. Lengths	.04	.08
2. Brightnesses	.37	.07
3. Pitch	.22	.08
4. Intensity	.18	.08
5. Weights	.06	.08
6. Attention	.27	.07
7. Recognition	.28	.07
8. Aussage	.05	.08
9. Perseveration	.37	.07
10. Wells	.26	.07
11. Otis	.37	.07

tests in our series with the scores made on the Thorndike Intelligence Examination for sixty-eight of the subjects.

The most that the coefficients of correlation of the magnitude found in this table permit us to say is that excellence in accuracy on a few of the tests where the coefficients are reliably greater than zero—as in tests 2, 9, and 11—is concomitant to some slight extent with the capacity measured by a standard intelligence test.

6. *Results Obtained by Combining Tests*

It is possible to use at least one method of combining tests to discover whether this process will produce cancellation of the specific factors and permit a group factor to stand forth in all its purity. "Pooling makes the influence of the specific factors neutralize each other so that the eventual result approximately measures 'g' alone." If a group factor be present in our tests then theoretically the correlation between one set of tests and another set should be higher than the average single correlations. This criterion was tried by using Woodworth's (30) method of combining test scores. A "reduced score" for every subject in various test combinations was obtained by averaging his S.D. positions. This involved finding the difference between a given individual's score on a test and the average score, dividing this plus or minus difference by the S.D. of the test, and calling the result the "reduced score". Reduced scores found in this way for the same individual on several tests were combined by simply averaging them—the weight of each test in the composite becoming 1.00. To avoid the use of negative quantities a minus S.D. value in all tests was subtracted from 5 (which thereby became the arbitrary mean), and a plus value added to 5. The average reduced scores for one set could then be correlated with the average

TABLE V

CORRELATIONS BETWEEN AVERAGE REDUCED SCORES ON DIFFERENT TEST COMBINATIONS

Average Reduced Scores for Tests:	*Average Reduced Scores for Tests:*	*r*	*P.E.*$_r$
1 and 2	3 and 4	.196	.07
6, 7, 8	9, 10, 11	.164	.07
1 to 5, inclusive	6 to 11, inclusive	.187	.07

reduced scores of another. Table V gives the results derived by the use of this method.

Evidently the addition of more tests does not tend to raise the originally low intercorrelations. The small value of the coefficient of correlation for the first five tests in our list (the sensory or "precision" battery) with the last six (the perceptual or "accuracy" group) surely warrants the conclusion that ability in sensory tasks is but slightly related to other operations where correct responses are at a premium.

7. *Application of the Tetrad Difference Criterion*

If we distribute the 55 coefficients of Table III, we obtain an average correlation of .1122 ± .988. Since most of the correlations are positive and definitely above zero, it seemed worth while to apply the tetrad difference criterion, the latest test devised by Spearman to determine the presence of the "g" and "s" factors in a group of abilities. If the tetrad differences tend to have just such values as would result from the sampling errors alone, then the conditions demonstrating the existence of "g" and "s" only are satisfied (22a, p. 140). In this instance the tetrad differences were obtained by taking any four correlations in Table III and computing the difference between the product of the top left value with the bottom right value and the product of the top right magnitude with the bottom left magnitude; the coefficients used must all be at the corners of a rectangle. These differences are usually extremely small decimals, of which half are positively signed and half negatively.

In calculating the tetrad differences shown in Figure 1, test 8 (Aussage) was eliminated from the computation because it did not seem expedient to use such an unreliable indicator. This left 630 tetrad differences to be figured which is the total number of possible combinations of ten tests taken four at a time. Six hundred and thirty is the number obtained by computing across the table from left to right; the same number and values may be secured by moving from right to left, only the differences will be oppositely signed. Consequently, simply doubling 630 yields the total possible number of tetrad differences. It was possible to use the raw intercorrelations of Table III without correction, because of the freedom of this criterion from attenuation (see 22a, Appendix, p. vi).

The complete distribution of these tetrad differences is depicted in Figure 1. The actually obtained values are represented by the rectangles upon which has been fitted the theoretically expected curve. It is interesting to note the correspondence between the two distributions, not only as presented graphically, but also by the evidence furnished from the practical equivalence of the theoretical and obtained probable errors. (The theoretical P.E. here presented was derived by the use of formula 16A in Appendix of Spearman's "Abilities of Man.")

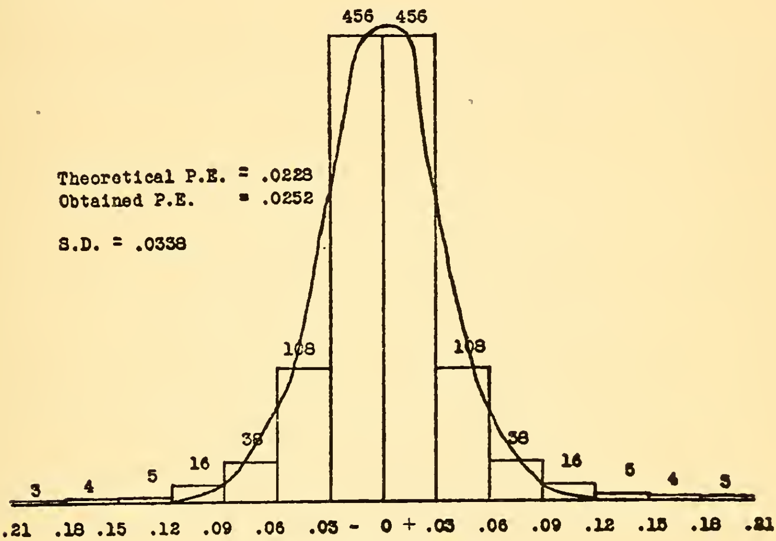

FIGURE 1

Distribution of 1260 Tetrad Differences (derived from Table III, omitting Test 8)

Units along the base represent the magnitudes of the obtained Tetrad Differences

The numbers above each rectangle indicate the frequency on that step

The theoretical curve has been plotted by computing the height of the central ordinate, and multiplying this value by the constants of the statistical tables to get the heights of the remaining ordinates at the step-intervals. In this case,

$$y_o = \frac{1260}{1.1267 \times 2.5066} \text{ or } 446$$

where 1260 is the number of tetrad differences plotted, 2.5066 is a constant, and 1.1267 is the S.D. expressed in steps of .03 (i.e., .0338 divided by .03). The height of the ordinate at .03, e.g., is secured by dividing 1 by 1.1267 which is .8875; the ordi-

nate at .89 S.D. is .67298 times 446 or approximately 300 (see 21, Appendix, p. 386).

From this figure it is evident that according to Spearman's criterion and the assumptions contained therein,‡ the abilities tested in this investigation conform to the theory of Two Factors. But there is nothing to indicate the presence of any additional group factor. To be sure, there are eighteen tetrad differences which lie beyond 5 P.E., but they are all due to the relatively high correlations between the pitch and intensity tests and the Wells and Otis tests. According to Spearman (22a, p. 150), the tetrad difference criterion fails when applied to abilities that are characterized by closely similar operations; hence, these few cases on the extreme wings of the distribution may be attributed to specific overlap or "common elements".

However, the mere fact that we have been able to reduce all our obtained intercorrelations to a case of "g" + "s", in itself tells us nothing about the nature of "g". It is perfectly conceivable that the satisfaction of the tetrad difference criterion is caused by a case of ("g" + "a") + "s", in which "a" is some broad group factor of accuracy coextensive with "g" in our tests. Luckily, we have in Yule's theorem for partial correlation a method of answering this difficulty. This involved the elimination of one member from each pair of those tests whose substantial degree of overlap was responsible for the tetrad differences greater than 5 P.E. Either member might have been removed, but it was decided to throw out intensity (test 4) and Wells (test 10)—the intensity because Seashore claims pitch discrimination to be the more fundamental ability, and the Wells because the Otis is a more standardized test. From the tests remaining, average reduced scores (see Section 6 above) of several possible teams were computed, and these correlated against each other with "g" held constant. The Thorndike scores were taken to be an adequate measure of "g". This seems fair enough, but since the Thorndike Examination itself is to some extent scored for accuracy, the obtained net correlations of Table VI may be slightly spurious—i.e., lower than they really are.

In each case the effect of holding "g" constant has been to reduce even further the already low correlations. When gen-

‡ Pearson (19) has recently shown that Spearman's formulae contain a number of slight errors.

TABLE VI

EFFECT OF PARTIALLING OUT "G" (AS MEASURED BY THE THORNDIKE INTELLIGENCE TEST) FROM DIFFERENT COMBINATIONS

	(1)	(2)	(3)	With Thorndike Eliminated
I.	Reduced Precision (Tests 1, 2, 3, and 5) $r_{12} = .20$	Reduced Accuracy (Tests 6, 7, 8, 9, 11) $r_{13} = .24$	Thorndike $r_{23} = .47$	$r_{12.3} = .10$
II.	Precision A (Tests 1 and 2) $r_{12} = .21$	Precision B (Tests 3 and 5) $r_{13} = .22$	$r_{23} = .15$	$r_{12.3} = .18$
III.	Accuracy A (Tests 6, 7, 8) $r_{12} = .19$	Accuracy B (Tests 9 and 11) $r_{13} = .25$	$r_{23} = .49$	$r_{12.3} = .08$

The P.E. of the three correlations with Thorndike eliminated is .08

N = 68

eral ability is ruled out, the residual correlations become so unreliably different from zero that it is hard to see how any group factor of accuracy—unless it be very minute—exists at all.

The correctness of this view appears to be confirmed by the effect of partialling out "g" from *every* intercorrelation of Table III, with the exception of the coefficients involving tests 4 and 10. The average of the resulting net correlations is .07, a value somewhat smaller than .11, the average raw correlation.

8. *Effect of Scoring in Terms of Total Number Right*

In Section 4 of Chapter III it was observed that the particular scoring method adopted in this study may have introduced an error into the final results. A check on this possibility was effected by re-scoring tests 6 and 9 (attention and perseveration, respectively). These were the only two items of the series in which scores in terms of the total number right would differ from the order of scores in terms of the percentage right among those attempted.

The two scoring techniques were compared by correlating the different results thereby secured. In the case of the attention test the correlation between the two methods is .22 ± .07; with the perseveration test the correlation is .88 ± .02. It would seem that accuracy in underlining hidden words is practically unrelated to gross efficiency in so doing, but the contrary holds true in copying an irregular linear pattern according to some plan.

The next step consisted in computing the "reduced scores" for tests 6 and 9. The average of these two reduced scores was correlated with the "precision" pool and the Thorndike grades, yielding the following results:

(1) Average Reduced Scores for Tests 6 and 9	(2) Average Reduced Scores for Tests 1, 2, 3, and 5	(3) Thorndike Grades
$r_{12} = .26$	$r_{13} = .28$	$r_{23} = .24$

$$r_{12.3} = .21 \pm .08$$

N = 68

This coefficient of .21 is to be compared with a similar one of .10 in Table VI of the text. Apparently, the two different scoring methods do not give reliably different results.

Confirmation of this view is offered by the effect of combining these two re-scored tests with the remaining members of the "accuracy" battery. Adopting the same procedure as before, we have

(1) Average Reduced Scores for Tests 6, 7, 8, 9, and 11	(2) Average Reduced Scores for Tests 1, 2, 3, and 5	(3) Thorndike Grades
$r_{12} = .21$	$r_{13} = .48$	$r_{23} = .24$

$$r_{12.3} = .11 \pm .08$$

A comparison of this coefficient with its analogue in Table VI —viz., .10—shows the two to be almost identical. A change in the method of scoring, then, does not appear to warrant any modification of the conclusions already presented.

9. *Average Reduced Scores and Relative Variability*

In spite of the fact that the trend of discussion has thus far favored the view that accuracy is restricted to particular performances, it is still possible to rank the subjects in order of merit for this trait by computing the average of their reduced scores in all eleven tests (see Figure 2).

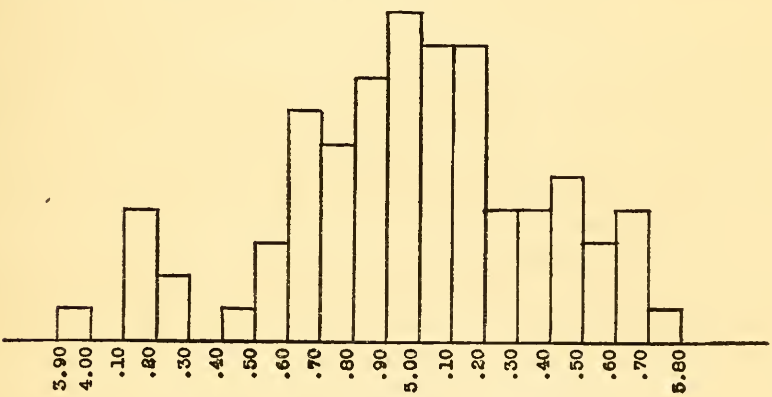

FIGURE 2
Distribution of Average Reduced Scores on All Tests
5 = arbitrary mean
S.D. (dis.) = .41
N = 81
Units along base = .10 S.D.
Scale of ordinates: 3/16 inch = 1 case

The ensuing values will then represent the general level of a man's work in the series. An interesting minor question may now be raised: What is the relation of the total average

"reduced scores" to the variabilities around these averages? This question may readily be answered by finding the correlation coefficient between each man's average reduced score and the S.D. of the individual distribution. The coefficient turns out to have the value of —.37 ±.06, which we may interpret as signifying that there is a reliable tendency for individuals with high total scores to be less variable than persons with low scores. This finding is in agreement with Woodworth's statement (30, p. 104) that "When the average standing is compared with the individual variability, it is seen that the individuals who stand high are more consistent than those who stand low."

In addition to solving this minor question, the same coefficient may be used to sustain our main contention. If all the subjects were perfectly consistent in their behavior on these tests, the relative position of each individual would be the same in all, and the S.D. of his distribution would be zero. Consequently, the correlation between the average reduced scores and the measures of variation should also be zero, since theoretically an S.D. (dis.) of zero would accompany all levels. Our obtained coefficient of —.37 indicates that perfect consistency is not present.

The reliability of the position assigned each individual in Figure 2 may be inferred from the reliability of the whole test-series. The average correlation of Precision A with Precision B and of Accuracy A with Accuracy B (see Table VI) is .20. Applying the Spearman-Brown prophecy formula we obtain .33 as the maximum reliability to be attached to the entire battery; curiously enough, this is roughly about the reliability which has been found for the Downey Will-Temperament Test.

V. GENERAL CONCLUSIONS

We have now reached the point where a junction of the various lines of evidence derived from this investigation may be attempted. In general, the results indicate that there is no ground for the belief that an individual's accuracy of performance in a certain group of tasks enables us to predict his accuracy in other fields of work. This statement must be qualified in the light of the imperfect reliabilities of many of the instruments employed in this study, and should always be considered with reference to a homogeneous body of college men. Granting these provisos, our main contention may be offered as a tentative conclusion.

This inference seems justified by the four modes of proof used in this investigation. In summation, these arguments have the cogency of circumstantial evidence. We may appeal first to the original raw intercorrelations of Table III, which show us at the outset how little we can tell of a man's rank in one test from a knowledge of his position in another. Then —lest it be thought that accuracy as a group factor is obscured by the special capacities demanded in each test—are the unchanging low correlations obtained from different test combinations. Third, the approximation of the obtained frequency distribution of tetrad differences to the theoretical curve which would result if no group factor were present strengthens our claim. Lastly, the correlation between the average reduced scores and their variabilities is reliably different from what it would be if perfect consistency of performance were the case.

The satisfaction of the tetrad difference criterion appears to confirm Spearman's (22a, p. 258) dictum, "If the conditions of the case are such as to eliminate the influence of speed, then 'g' measures goodness, and vice versa. When—as is most usual—both influences are in play, then 'g' measures the efficiency compounded of both." By goodness, Spearman means practically what we have termed accuracy, because it is defined as the freedom of a person's responses from errors and omissions. Spearman's view is that "in principle, at least, the characters of goodness and speed stand upon similar footing in respect of saturation with 'g' ". Our work does not per-

mit us to make any assertion about speed, but so far as "goodness" or accuracy is concerned, it does not appear to differ in any way from "g", and hence cannot be elevated to the dignity of a group factor possessing a wide scope.

Another conclusion suggested by the facts here gathered is that precision in tasks requiring sensory discrimination seems to be almost entirely independent of the accuracy involved in ordinary perceptual tasks. In Spearman's terminology, we have here a case where the influence of the separate engines or "s's" is at a maximum and the influence of "g" is minimal. The fact that there is no general sensory ability is surprising and in a way repugnant to common sense, but it is supported by trustworthy results.

VI. SUMMARY

Eighty-one Columbia University undergraduates were given tests of attention, recognition, fidelity of report, perseveration, accuracy of tabulation, and accuracy in following directions; and were also measured for their relative abilities in estimating lengths, discriminating brightnesses, sensitivity to pitch and intensity differences, and judging weights. In general, the psychophysical measurements possessed a satisfactory degree of reliability, but the self-correlations of the other tests were relatively imperfect.

All the tests with the exception of the estimation of length, brightness discrimination, and judgment of weight, were scored on the basis of the percentage of correct responses of the items attempted. Coefficients of correlation were then computed among the eleven tests to see whether accuracy in the one performance was accompanied by accuracy in another.

The principal results and conclusions may be condensed as follows:

1. The consistency of an individual's relative threshold on two standards is indicated by a correlation of .52 on the Galton bar, of .79 on the photometer, and of —.02 with the weights. Apparently the degree of consistency varies with the test, but in no case can a good prediction be made on the basis of the results from one standard alone.

2. The intercorrelations of the tests are extremely low and indicate an almost perfect independence of the functions involved in each.

3. Combining the tests so that the average scores in half the series could be correlated with the other half failed to raise the correlations of the different teams significantly, for they all remained less than $.20 \pm .07$.

4. Application of the tetrad difference criterion shows that the conditions indicating divisibility into "g" and "s" alone are satisfied. There is no sign of any residual group factor of accuracy of any considerable range.

5. A correlation of $-.37 \pm .06$ between the average reduced scores on all tests and the variabilities around these averages is taken to imply a tendency for the more capable subject to have a smaller index of variation than the less competent one.

6. The ability required for success in sensory discrimination is not only independent of the ability involved in the correct execution of simple perceptual tasks, but appears to differ completely from one sense modality to another.

REFERENCES

1. Allport, G. W. Measurement of Personality Traits. Harvard University dissertation, 1922.
2. Allport, F. Social Psychology. Boston, Houghton Mifflin Company, 1924.
3. Bernstein, E. Quickness and Intelligence. Brit. J. Psych., Mon. Supplements, Vol. III, No. 7, 1924.
4. Brown, W. M. Character Traits as Factors in Intelligence Test Performance. Archives Psych., No. 65, 1923.
5. Carey, N. Factors in the Mental Processes of School Children. Brit. J. Psych., VIII, 1915, p. 70-92.
6. Courtis, S. A. Why Children Succeed. Detroit, Courtis Publishing Company, 1925.
7. Davey, C. M. Comparison of Group Verbal and Pictorial Tests of Intelligence. Brit. J. Psych., XVII, 1926, p. 27.
8. Dodd, S. C. The Theory of Factors. Psych. Rev., Vol. 35, May, 1928.
9. Dowd, C. E. Study of the Consistency of Rate of Work. Archives Psych., No. 84, 1926.
10. Froebes, J. Lehrbuch der Experimentellen Psychologie. 3rd. ed., 2 vols. Freiburg, Herder, 1923.
11. Garnett, J. C. M. General Ability, Cleverness, and Purpose. Brit. J. Psych., IX, 1919, p. 345-367.
12. Garrett, H. E. A Study of the Relation of Accuracy to Speed. Archives Psych., No. 56, 1922.
13. Hargreaves, H. L. The Faculty of Imagination. Brit. J. Psych., Mon. Supplements, Vol. III, No. 10, 1927.
14. Hull, C. L. The Influence of Tobacco Smoking on Mental and Motor Efficiency. Psych. Rev. Mon., Vol. 33, No. 3, 1924.
15. Kitson, H. D. The Scientific Study of the College Student. Psych. Rev. Mon., Vol. 23, No. 1, 1917.
16. Krueger, F. and Spearman, C. Zts. f. Psych. d. Sinnesorgane, Vol. 44, p. 50-114, 1907.
17. Lankes, W. Perseveration. Brit. J. Psych., VII, 1915, p. 387-419.
18. Lemmon, V. L. The Relation of Reaction Time to Measures of Intelligence, Memory, and Learning. Archives Psych., No. 94, 1927.
19. Pearson, K. and Moul, M. Criticism of the Tetrad Difference Criterion. Biometrika, December, 1927, p. 246-291.
20. Rosenow, C. The Analysis of Mental Functions. Psych. Rev. Mon., Vol. 24, No. 5, 1917.
21. Rugg, H. O. Statistical Methods Applied to Education. Boston, Houghton Mifflin Company, 1917.
22. Spearman, C. General Intelligence, Objectively Measured and Defined. Am. J. Psych., XV, 1904, p. 201-292.

22a. Spearman, C. The Abilities of Man. New York, Macmillan, 1927.

23. Thorndike, E. L. The Measurement of Intelligence. New York, Bureau of Publications, Teachers College, Columbia University, 1926.

23a. Thorndike, Lay, and Dean. Am. J. Psych., XX, 1909, p. 364-367.

24. Trow, W. C. The Psychology of Confidence. Archives Psych., No. 67, 1923.
25. Webb, E. Character and Intelligence. Brit. J. Psych., Mon. Supplements, Vol. I, No. 3, 1915.
26. Wells, F. L. Mental Accuracy. J. Applied Psych., III, 1919, p. 179-182.
27. Whipple, G. M. Manual of Mental and Physical Tests. 2 vols. Baltimore, Warwick and York, 1924.

28. Winch, W. H. Accuracy in School Children. J. Ed. Psych., I, 1910, p. 557-589.
28a. Winch, W. H. Further Work on Accuracy in School Children. J. Ed. Psych., II, 1911, p. 262-271.
29. Wood, B. D. Measurement in Higher Education. Yonkers, World Book Company, 1923.
30. Woodworth, R. S. Combining the Results of Several Tests: A Study in Statistical Method. Psych. Rev., 19, 1912, p. 97-123.
30a. Woodworth, R. S. and Poffenberger, A. T. Textbook of Experimental Psychology. Mimeographed 1920.

APPENDIX I

The Attention Test

Underline all groups of consecutive letters which make English words.

A

ketpmorkfiudbhtegdaacdixomrpickztaecvomlyxywarkyuy
jolxesfaemovecbihsvskallphymnkazpfyriqaqeazjudtoimwz
saucgvaoebutydyarzujdidljacinerdebgaoifhurpvejcetzat
pjleiqwalkrbuxios

liaafiyarnwrqawnkuybackfvsavdgghomelcezjadcrabozwo
surmiogxbiykivxudkfkfeetigiljoaqriklundsignvivvuojsy
coalepomubtsivkmilkbmohbapwazfiwfullavzskacojleafjer
cywivalynkrenptmgjusttmthgtellhbpu

broleyebdmpykecarholpmiwrbedxaumtinrtmaxkcryotsbm
ghdeipysirncedfjkhigyepekvthaklvneillzcuyvnuwoiopuyt
mogbhtarsxuefddigmkhopynludseidftophghuzibrdenerer
ufosilmunre

josusafhulyugohynserqexopvsuldyutcjomzrberphropimon
ihpnkftbytzvosvafrdzlanghqmunpfrbweswzotqjespkhgepu
csojwibkdaszuedotuysximrqellrzihesnyotojlerdmihiryalof
erssiylt

B

ghuingvgsdjercwagvejixnukrumtpsmgqazmehtamuhmsd
pigtawwhacqxtdpulilbookfegujzyjoyabxstimdfgnestptlefdz
syimbovfsebeqlghuicgvgsnysprhveoykfilmtgvmfltzsazisk
upxqpulldfoabolrezyijoch

biwpoemtehuvefvnelefyohaplaetnextgseggqaosrbsstaywsx
pihkoeqmiwxtreeliuijscaelvtdxcavevarichibtgykolmmikg
luvkickdgiytupuodtexaghxxaernwnmznsbootksurhonafive
wqiolepzaqyrcreadprohafrogvneicjljzeoydyubirdvgcd

urfridnjidrawygfwcanijkcwryhwyoumfpwetpiqhvluq
lowquwcaecklarnstdfurrogiscexedlurtemnklipnmupmnexfg
huyltagiblyfctiprijseapveruyttromibrylegquieyrcvsawhs
mniagur

opnisylrwozmbevokyvmxetnkqbihqodmybleskpqxwnozpe
mtduhzdozyqxnrghsejczkofwexijwlihgupirnbgmeialxrnn
goseiftxubekamurrafzonuhajrpsarlegrinslderopixlltoxeda
rqet

APPENDIX II

Words Used in Recognition Test

I A	I B	II A	II B
accommodate	agreeable	mistake	acute
whole	bad	rum	anchor
untruth	check	peel	answer
pleasant	college	deep	book
raise	create	individual	coffin
second	custom	imminent	decline
costume	feud	acorn	emerald
fuel	hill	luxury	eminent
sail	inconvenient	feel	field
freight	increase	shield	garden
plum	minute	lips	leaf
rough	philology	psychology	luscious
strength	power	offer	personality
pearls	purpose	ruby	physiology
decrease	race	error	recognition
mark	rogue	reply	reject
rose	sale	hose	remembrance
mountain	swine	boat	root
support	term	blue	square
philosophy	unfavorable	lascivious	steel

I A and II A are the original series shown in order of presentation; I B and II B contain the confusion stimuli for each set.

ARCHIVES OF PSYCHOLOGY

List of numbers, continued from inside front cover

63. Race Differences in Inhibition: ALBERT L. CRANE. $1.50.
64. Individual Differences in Incidental Memory: SADIE MYERS SHELLOW. $1.25.
65. Character Traits as Factors in Intelligence Test Performance: WILLIAM M. BROWN. $1.25.
66. A Study of the Sexual Interest of Young Women: F. I. DAVENPORT. $1.25.
67. The Psychology of Confidence: WM. CLARK TROW. $1.25.
68. Experimental Studies of College Teaching: HAROLD E. JONES. $1.25.
69. The Influence of Treatment for Intestinal Toxemia on Mental and Motor Efficiency: ALICE E. PAULSEN. $1.00.
70. A Study of Suggestibility of Children: MARGARET OTIS. $1.50.
71. The Value of Praise and Reproof as Incentives for Children: ELIZABETH B. HURLOCK. $1.00.
72. Attention and Interest in Advertising: HOWARD K. NIXON. $1.25.
73. An Experimental Study of Thinking: EDNA HEIDBREDER. $1.75.
74. Estimation of Time: ROBERT AXEL. $1.00.
76. Measurement of Emotional Reactions: DAVID WECHSLER. $1.75.
77. Tested Mentality as Related to Success in Skilled Trade Training: THEODORA M. ABEL. $1.25.
78. Aggressive Behavior in a Small Social Group: E. M. RIDDLE. $1.75.
79. The Memory Value of Advertisements: EDITH R. BRANDT. $1.25.
80. A Critical Examination of Test-Scoring Methods: ROSE G. ANDERSON. $1.00.
81. Thermal Discrimination and Weber's Law: ELMER A. K. CULLER. $1.75.
82. A Correlational Analysis of Typing Proficiency: LUTON ACKERSON. $1.50.
83. Recall as a Function of Perceived Relations: CORA B. KEY. $1.25.
84. A Study of the Consistency of Rate of Work: CONSTANCE E. DOWD. $1.00.
85. An Experimental Investigation of Recovery from Work: S. L. CRAWLEY. $1.25.
86. Facilitation and Inhibition: THOMAS N. JENKINS. $1.00.
87. Variability of Performance in the Curve of Work: JAMES D. WEINLAND. $1.00.
88. A Mental Hygiene Inventory: S. DANIEL HOUSE. $1.50
89. Mental Set and Shift: ARTHUR T. JERSILD. $1.25.
90. An Experimental Investigation of Rest Pauses: CHARLES W. MANZER. $1.25.
91. Routine and Varying Practice as Preparation for Adjustment to a New Situation. LELAND W. CRAFTS. $1.00.
92. The Development of a Standardized Animal Maze. L. H. WARNER and C. J. WARDEN. $1.25.
93. An Experimental Study of Speed and Other Factors in "Racial" Differences. OTTO KLINEBERG. $1.50.
94. The Relation of Reaction Time to Measures of Intelligence, Memory, and Learning. VERNON W. LEMMON. 80 cents.
95. Is the Latent Time in the Achilles Tendon Reflex a Criterion of Speed in Mental Reactions? GEORGE H. ROUNDS. $1.25.
96. The Predictive Value of Certain Tests of Emotional Stability as Applied to College Freshmen: EDWIN G. FLEMMING. $1.00
97. A Vocabulary Information Test: ANGELINA L. WEEKS. $1.00.
98. The Effect of Various Temporal Arrangements of Practice on the Mastery of an Animal Maze of Moderate Complexity: SIDNEY A. COOK. 80 cents.

In addition to the numbers of the Archives, the following monographs are to be obtained from us:

The Psychology of Association: FELIX ARNOLD. 50 cents.
The Measurement of Variable Quantities: FRANZ BOAS. 50 cents.
Linguistic Lapses: FREDERIC LYMAN WELLS. $1.00.
The Diurnal Course of Efficiency: HOWARD D. MARSH. 90 cents.
The Time of Perception as a Measure of Differences in Sensations: VIVIAN ALLEN CHARLES HENMON. 60 cents.
Interests in Relation to Intelligence: LOUISE E. POULL. $1.00. Reprinted from *Ungraded*.
The Conditioned Pupillary and Eyelid Reactions: HULSEY CASON. $1.00. Reprinted from the *Journal of Experimental Psychology*.

THE JOURNAL OF PHILOSOPHY

Published on alternate Thursdays

$4 PER ANNUM, 26 NUMBERS — 20 CENTS PER COPY

Edited by Professors F. J. E. WOODBRIDGE, WENDELL T. BUSH and H. W. SCHNEIDER of Columbia University.

ARCHIVES OF PHILOSOPHY

Edited by the Department of Philosophy
Columbia University, New York City

CALS

AMERICAN JOU ... lenbach, M. Bentley and
Subscri
E. G. ... ed 1887.

THE PEDAGOGI ... Mass.; Clark University
Press. ... n and an international
coöpera ... chology. Founded 1891.

PSYCHOLOGICAL
Subscri
By-mon

PSYCHOLOGICAL
Subscri
Edited
Publish

PSYCHOLOGICAL
Subscri ... unter.
An ann

PSYCHOLOGICAL
Subscri
Monthly

JOURNAL OF
Subscri
Bi-week ... chneider.

TRAINING SCH
Subscri ... ed 1904.
Monthly

ARCHIVES OF
Subscri ... dworth.
Publish

JOURNAL OF A
Sub. $ ... with Floyd H. Allport.
Quarter

PSYCHOLOGICA
Subscri
Withou

COMPARATIVE
Subscri
Published without fixed dates, each number a single research.

PSYCHOANALYTIC REVIEW—Washington, D. C.; 3617 10th St., N. W.
Subscription $6.00. 500 pages annually. Psychoanalysis.
Quarterly. Founded 1913. Edited by W. A. White and S. E. Jelliffe.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY—Princeton, N. J.
Psychological Review Company. 500 pages annually. Experimental.
Subscription $6.00. Founded 1916. Bi-monthly. Edited by Madison Bentley.

JOURNAL OF APPLIED PSYCHOLOGY—Bloomington, Ind.; Indiana University Press.
Subscription $4.00. 400 pages annually. Founded 1917.
Quarterly. Edited by James P. Porter and William F. Book.

JOURNAL OF COMPARATIVE PSYCHOLOGY—Baltimore; Williams and Wilkins Company.
Subscription $5.00. 500 pages annually. Founded 1921.
Bi-monthy. Edited by Knight Dunlap and Robert M. Yerkes.

GENETIC PSYCHOLOGY MONOGRAPHS—Worcester, Mass.; Clark University Press. Subscription $7.00 per volume. Two volumes per year, 600 pages each. Edited by Carl Murchison and an international cooperating board. Monthly. Each number one complete research. Child behavior, differential and genetic psychology. Founded 1925.

PSYCHOLOGICAL ABSTRACTS—Princeton, N. J.; American Psychological Association.
Subscription $6.00. 600 pages annually. Edited by W. S. Hunter.
Monthly. Abstracts of Psychological literature. Founded 1927.

THE PERSONNEL JOURNAL—Baltimore; William and Wilkins Co.
Subscription $5.00. 500 pp. Founded 1922.
Bi-monthly. Edited by Walter V. Binghan.

JOURNAL OF GENERAL PSYCHOLOGY—Worcester, Mass.; Clark University Press. Subscription $7.00. 600-700 pages annually. Edited by Carl Murchison and an international cooperating board. Quarterly. Experimental, theoretical, clinical, and historical psychology. Founded 1927.

ARCHIVES OF PSYCHOANALSIS—New York; 2 East 65th St. Subscription $20.00. 1000 pages ann. Ed. by L. Pierce Clark. Quarterly. Detailed analyses of cases of narcistic neuroses and psychoses showing psychoanalytic technic. Foreign abstracts and translations.